medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted August 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Title: User testing of a Diagnostic Decision Support System with Automated Chart Review to Facilitate Clinical Genomic Diagnosis

Authors

Alanna Kulchak Rahm[1]

Lynn K. Feldman[2]

Makenzie A. Woltz[1]

Thomas N. Person[1]

Jonathon C. Reynolds[1]

Troy Jenkins[3]

Conner Jenkins[3]

Joseph Peterson[3]

Peter N. Robinson[4]

Marc S. Williams[1]

Nephi A. Walton[5]

Michael M. Segal[2]

Corresponding author

Name: Alanna Kulchak Rahm

Address: Genomics Medicine Institute, Geisinger Health System, 100 North Academy Avenue, Danville PA 17822-2620, USA

Email: akrahm@geisinger.edu

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted August 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Phone: (570) 214-5093

AUTHOR AFFILIATIONS

1 Genomic Medicine Institute, Geisinger, Danville, Pennsylvania, USA

2 SimulConsult Inc., Chestnut Hill MA, USA

3 University of Utah, Salt Lake City, UT, USA

4 The Jackson Laboratory for Genomic Medicine and Institute for Systems Genomics, University of Connecticut, Farmington, CT, USA.

5 Intermountain Precision Genomics, Intermountain Healthcare, St George, Utah, USA

Keywords: health information systems, heuristic evaluation, cognitive walkthrough, Diagnostic Decision Support System (DDSS), Clinical Decision Support (CDS),

Word count: 3573

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted August 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

ABSTRACT

Objective: There is a need in clinical genomics for systems that assist in clinical diagnosis, analysis of genomic information and periodic re-analysis of results, and can utilize information from in the electronic health record to do so. Such systems should be designed using the concepts of human-centered design, fit within clinical workflows, and provide solutions to priority problems.

Materials and methods: We adapted a commercially available diagnostic decision support system (DDSS) to use extracted findings from a patient record and combine them with genomic variant information in the DDSS interface. Three simulated patient cases were created, and a structured interview guide was used to assess technical issues, workflow fit, and acceptability. Three testers conducting evaluation of genetic conditions were selected for user testing; 3 additional testers were included to evaluate fit and system implementation potential. Qualitative analysis of test session transcripts was done.

Results: Tester responses were positive and noted good fit within real-world clinical genetics workflow. Technical issues related to interface, interaction, and design were minor and fixable. Testers suggested solving issues related to terminology and usability through training and added infobuttons. Time savings was estimated at 30-50% and other uses for in-house clinical variant analysis were noted to increase fit with workflow and address priority problems.

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted August 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Discussion and Conclusion: We provide preliminary evidence that a system including DDSS and automated chart review may help improve efficiency and consistency of genetic diagnosis and addresses perceived needs for clinical end-users. Further development could improve functionality and implementation potential.

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted August 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

LAY SUMMARY

There is a need in clinical genomics for tools that assist in analysis of genomic information and can utilize information from the electronic health record to do so. Such tools should be usable, fit within clinical workflows, and provide solutions to priority problems. We adapted a commercially available diagnostic decision support system (DDSS) to use findings taken from a patient record. The DDSS then combined the patient record findings with genomic data and presented this to clinicians to help evaluate the patient. Three clinicians who routinely diagnose genetic conditions and three other clinicians who also use genomic information tested the DDSS. Tester responses were positive and noted good fit within the clinical genetics workflow. Technical issues identified were minor and fixable. Testers suggested solving other issues through training and clickable links. Time savings was estimated at 30-50% and other uses for in-house clinical genetics were noted to increase fit with workflow and address other priority problems in genetics. This study provides preliminary evidence that a DDSS with automated chart review may help improve the genetic diagnosis process and addresses perceived needs for clinical end-users. Further development could improve functionality and implementation potential.

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted August 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

BACKGROUND AND SIGNIFICANCE

Clinical Decision Support (**CDS**) integrated into Electronic Health Records (**EHR**s) has long been considered a promising way to improve patient outcomes and decrease inefficiencies.[1-4] It is recognized that CDS must be designed with the user in mind, fitting the concepts of human-centered design with computer interfaces at the individual clinician level.[1,5] To facilitate implementation, CDS must also fit within clinician workflow and provide a solution to a priority problem for the clinician and the healthcare system.[4,6-8]

Diagnostic Decision Support Systems (**DDSS**s) are a key type of CDS needed in genomics to supplement a shortage of trained clinicians and address the inherent complexity of genomic diagnosis.[9,10] Complexity arises from the heterogeneous nature of genetic diseases, the variable expression in patients, and the degree of overlap in findings (i.e., signs, symptoms and test results) among genetic conditions, sometimes differentiated mainly by onset age of individual findings.[11]

Systematic review and position statements note two needs from DDSSs in genomics: (1) a cost-effective, regular approach to re-evaluation of patient cases in light of new findings or genetic knowledge, when testing does not immediately yield a diagnosis; and (2) developing automation of chart review.[12,13] Most genomic patient EHR files are extensive and written by multiple clinicians such that manual review is prohibitively time-consuming, increasing the risk of missed information that could facilitate timely diagnosis and avoidable costs from repeated or unnecessary tests. In addition, most relevant information is contained in clinical notes that are unstructured and requires

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted August 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

approaches such as natural language processing (NLP) to reduce the need for manual review.

Here, we report on user testing of an adaptation of an existing DDSS that addresses both needs. The goal was to study human-computer interaction, usability, and perceived fit with clinical need and workflow, and to assess the potential for implementation into the real-world clinical environment.

MATERIALS AND METHODS

Adapting a DDSS for automated chart review of clinical findings

We adapted an existing DDSS to automate chart review by accepting findings from an EHR. The DDSS was the SimulConsult Genome-Phenome Analyzer, a commercial product in clinical use that has been shown to be accurate and helpful in clinical diagnosis, including interpreting genomic results.[14-16] Clinicians enter findings and the DDSS returns a differential diagnosis (ranked list of candidate diseases). Also returned are suggestions of other findings to check, ranked by their usefulness in narrowing the differential diagnosis in a way that takes into account cost and treatability; thus facilitating the iterative approach of information gathering in diagnosis.[17,18] For each finding, presence (with onset age) or absence can be specified. The DDSS supports importation of annotated variant call files (**VCF**), and generates many outputs, including a Patient Summary for saving interim patient findings and a customizable genomic Return of Results (**RoR**) report that facilitates standardized communication for patients and referring clinicians and shown to be effective in previous testing.[19-22]

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted August 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

For **NLP** of EHR notes we used the cTAKES tool with the Unified Medical Language System (**UMLS**) module.[23] The DDSS was modified by (1) mapping its findings to Human Phenotype Ontology (**HPO**) and UMLS codes, including creation of hundreds of new HPO terms that then resulted in creation of new UMLS concepts, (2) using results from NLP analysis of EHR notes to flag mentions of the findings used by the DDSS, and (3) augmenting the DDSS's interface to present the flagged findings with contextual information needed to clinically assess the information. We designate this resulting prototype the Genotype-Phenotype Archiving and Communication System with SimulConsult (**GPACSS;** Figure A and Table 1). The philosophy behind GPACSS was to combine the automated search for findings using NLP with review by the clinician to minimize both false positives and false negatives in the following ways:

1. **Minimizing false negatives on NLP flagging of findings**. False negatives were reduced by including parent and child codes, e.g., the DDSS finding of intellectual disability also included codes for developmental delay and particular types of intellectual disability.
2. **Minimizing false positives through the DDSS Usefulness metric**: Flagged findings were displayed in the DDSS ranked using its usefulness algorithm,[24] de-prioritizing data of low relevance.
3. **Minimizing false positives through clinician verification**: Findings identified by NLP were indicated with a flag icon (Figure B); clicking the flag displayed the information needed by the clinician to assess reliability, presence versus absence, and onset. The displayed information included the various mentions in the EHR,

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted August 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

including date, observer identity, and 3 sentences: the one with the finding plus the preceding and subsequent sentences.

Table 1. Adaptations to build GPACSS

Adaptation	Component	Approach
Overall design	SMART-on-FHIR enabled **EHR**	• Logica platform (https://www.logicahalth.org/; formerly Health Services Platform Consortium; HSPC)
	Archive	• Custom archive stores key files • RESTful interface.
Coordination and communication	**User interface**	• SMART-on-FHIR application (GPACSS FHIR app client, Figure A). • Interface allows user access to DDSS directly from patient record. • Choice to launch with no findings or with findings previously saved
	Coordination	• GPACSS “Coordinator” API saves the NLP output • Matching of UMLS codes in NLP output to DDSS findings

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted August 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

		• Send the matched flagged findings to the DDSS at launch (Figure A)
Natural language processing	**Extraction of findings**	• NLP: open source Apache cTAKES 4.0 [23] • cTAKES default modules to handle sentence boundary detection, tokenization, normalization, tagging parts of speech, recognizing named entities, and negation. • cTAKES pre-trained module to recognize UMLS concepts in text
	Mapping in DDSS	• DDSS findings mapped within the DDSS to one or more Unified Medical Language System (UMLS) and Human Phenotype Ontology (HPO) codes • Mapping strategy minimizes false negatives in term capture while tolerating false positives (identifying information unrelated or irrelevant to the diagnostic process).
	Display in DDSS	• Findings identified by NLP display a flag icon • Clicking the flag enables viewing of metadata

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted August 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Creating simulated cases for user testing

Three cases of increasing complexity created for testing were real but de-identified (Table 2). The scenarios assumed that some patient characterization was already noted by the clinician and genomic results were now available and could be interpreted in light of clinical information in the EHR. For the 3 cases, a total of 5 findings were used for initial information before the genomic results, with 3 (one per case) being flagged findings identified through NLP.

Table 2. Creating and processing simulated patients

*3 simulated patients of escalating complexity with notes and clinical findings were created based on 3 real patients with known genetic disorders. Two others were used in earlier testing and the demo video (*https://simulconsult.com/videogpacss*).*

Simulation aspect	**Process used in simulation**
Phenotype	All findings discussed in the EHR text were retained but notes were manually rewritten by changing narrative to remove identifying information and prevent re-identification, shifting dates while maintaining chronological relationships, and shifting numeric values while maintaining appropriate range to retain clinical meaning (i.e. laboratory values for "high-abnormal" range were maintained in that value range).
Genotype	Genomic data was simulated by adding the known causal variant to variant tables generated from publicly available

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted August 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

	trios from the 1000 Genomes Project to prevent identification of a real person. However, the known causal variant was not given key annotations such as functional and conservation scores that are used in the DDSS's explanation of variant and zygosity severity scores.
Processing	Resulting simulated patient data for 3 simulated patients was loaded into the Logica platform for user testing and run through the cTAKES Clinical Pipeline using default settings.

User Testing Methods

Participants: All testers were staff members at Geisinger, where the Epic EHR has been in regular clinical use since 1996. Because GPACSS was designed to support differential diagnosis, 3 clinicians with primary roles involving differential diagnosis were purposively selected for user testing: a genetic counselor (pediatrics and research); a pediatric geneticist (orders exomes daily); and a general internal medicine physician (ordered 4-5 exomes in past month). In general, 3 - 5 evaluators are considered sufficient for this type of user testing.[25]. To further assess organizational implementation potential within a healthcare system through extended usability and fit of GPACSS, 2 additional testers from other areas of clinical genetics were purposively selected: a genomics laboratory director, and a genetic counselor involved in variant interpretation.

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted August 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Testers viewed a 4-minute training video demonstrating GPACSS use (https://simulconsult.com/research/flag), beginning from saved patient findings, importing a VCF, and additions of flagged findings to arrive at a diagnosis and create a customizable patient-friendly RoR report.

Testing Sessions: A semi-structured interview guide was created encompassing CDS design (information, interaction, interface) and implementation (acceptability, perceived need, feasibility, workflow fit) issues. An experienced interviewer (AKR) and observer (MAW) worked with each tester to imagine using GPACSS for each case. A “think aloud” approach with the interviewer asking questions as needed and at key points in the process resulted in a “cognitive walkthrough” and “heuristic evaluation.”[25,26] The user testing interview and process was piloted prior to use with a cancer genetic counselor in a one hour session and reviewing one test scenario. Data from the pilot tester was consistent with the additional testers selected for organizational implementation potential and thus is also included in the results. At the end of the session, testers were asked a series of questions to rate the overall usefulness of GPACSS and specific key components.

Transcripts were created from the audio portion of each user testing session. The computer screen was also video recorded during the session to capture tester movement through GPACSS. Testers were provided a $100 gift card. User testing protocol was reviewed and approved by the Geisinger IRB.

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted August 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Analysis

To analyze the user testing sessions, two coders viewed each user test session and read the transcripts. A codebook was created using themes identified across transcripts. Two coders organized the data in the transcripts according to codes (MAW, JCR); and thematic codes were evaluated and organized according to Nielsen's 10 Heuristics[27] and Miller et al,[1] which contains 42 CDS components across 3 categories (information, interface, and interaction). Rogers' Diffusion of Innovations in organizations constructs guided assessments for future implementation.[28]

RESULTS

Five of six testers completed 3 patient cases over a 2-hour test session. The pilot session included one scenario and lasted one hour. Results are presented from the 3 clinicians regularly performing differential diagnosis as this is the primary purpose of GPACSS. Results from the additional testers (n=3 including the pilot tester) related to future implementation potential are presented separately.

Performance of the GPACSS prototype

GPACSS performed all intended functions: automated chart review, interpretation of genomic variants to make clinical diagnoses, and interactive generating of an RoR report. The generic cTAKES NLP using the UMLS concepts found 20 of the 30 (67%) pertinent positive concepts that a pediatric neurologist (MMS) identified by reading the records. In all cases, this was sufficient to make the diagnoses.

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted August 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Usability of GPACSS prototype for differential diagnosis

Overall reception to the prototype was positive. Testers raised some general issues relevant to CDS design (information, interaction, interface); including terminology, functionality they had trouble finding, and the desire for more infobuttons and training.

Testers liked the flagged findings (Figure B), the contextual information for each mention in the EHR and the rank ordering of flagged findings. Testers also liked the visualization of the evolving differential diagnosis and the automated RoR report for sharing with patients and referring clinicians, including the ability to save and access this report from the EHR. Testers anticipated considerable time savings from use and wanted to know when it would be available in the clinic.

Testers were thoughtful and purposeful in use of GPACSS for the patient scenarios. Notably, in case 3 (the most complex case), one tester did not immediately choose the top diagnosis offered by GPACSS. Supported by the data displayed in the DDSS, to make a definitive diagnosis the tester indicated they would evaluate for the second-ranked disease – as that second condition had a test that was easy and accurate and the condition was also more treatable – indicating utilization of the DDSS as intended and consistent with clinical diagnostic decision-making.

*Interface***:** The interface was noted to be complex, but testers stated this was expected due to the inherent complexity of diagnosis, especially for genetic conditions. Users expected a learning curve to become proficient.

> *"It's going to take a lot to learn. A lot of clicking back and forth and it's not super intuitive but I get it" [Tester 2]*

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted August 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

"More training would be good… unless I was doing it all the time for all of my patients, every step, I might not realize that some of the features are available…" [Tester 3]

Placement, positioning, and multiple presentation layers (text and graphics in the interface)[1] were well liked. In particular, the "Assess diagnosis" display was seen as valuable, showing the logic used by the DDSS, by comparing patient findings to known information about the disease. Varying interpretations were noted by testers for the different graphical bars and shading used in the DDSS, however, this did not hinder their ability to use the information to make the diagnosis or to question the need for the bar itself. More labeling was suggested to help with interpretation.

"These bars are different lengths, so I assume it's having something to do with frequencies... so I'm not sure why this part is purple....if there were something [on the assess diagnosis tab] that said this is 100% over here and this is 0% over here, that would kind of help, if I knew that that was the case... I'm not sure what these other colors are referring to." [Tester 5]

"To me, the green bar in it shows me they are confident that this genetic variant aligns with the phenotypic markers that we have identified. I don't necessarily know how far the bars will tell me they're confidence in pathogenic versus VUS." [Tester 3]

Testers also expected some menu choices to be elsewhere than their current location, such as expecting the Genome RoR report to be under the DNA symbol rather than with

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted August 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

the rest of the report outputs. Some found the generic “share” navigation image confusing, saying it appeared to be an “exit” icon.

*Interaction***:** Testers initially expressed the opinion of “too many clicks” or brought up “click fatigue”, but subsequently noted where required clicking was unavoidable and necessary as the user testing session continued. For example, they expressed value in taking the time to correctly specify onset information (which requires clicking and cognitive load in the DDSS), as this is standard in the genetic diagnostic process and so were accepting of the additional clicks required for this task. Testers did comment on the cognitive load required to review and choose age of onset for flagged findings “Cognitive Load” in DDSS testing refers to additional thinking required of users to interact with the tool and the general recommendation is to minimize cognitive load when designing a CDS.[1] While testers initially noted it as an additional task, they ultimately did not identify any increase in cognitive load completing this task with the DDSS.

*Information***:** Testers appreciated existing resources, such as the hover feature that revealed synonyms to findings and requested even more infobuttons and hovers. Testers indicated confusion over some terminology, notably “zygosity” and “severity score” when reviewing the genomic variants. Some testers found the explanatory resource for the terms while others did not. The severity score however, might have been confusing as the annotation information explaining severity scores was not included in the VCF for the simulated patients.

The fact that the EHR mentions displayed in flagged findings were sometimes triggered by parent or child concepts was noticed by all testers, and some found the findings used

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted August 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

in the DDSS not at the level granularity they were expecting. Regardless, they emphasized the importance of being able to review the information to avoid false negatives in the NLP process.

*Acceptability and Fit for Differential Diagnosis***:** Testers saw immediate benefits from use through three key ratings:

- **Satisfaction**: average 8.5 out of 10, range 8-9.5, n=3;
- **Navigation:** average 8 out of 10, range 7.5 to 9, n=3; and
- **Time saved:** estimated "30-50%" [Tester 3].

These testers identified specific value from time saved for chart review and noted how the process fit with their clinical workflow diagnosing genetic conditions.

> *"Everything's there [in the chart] and the question is how easy is it to find. I'm sure if you're a malpractice lawyer you get very good at pulling stuff out of these charts and asking why didn't you see that. Yet I can't look at everything." [Tester 2]*

> *"This is stuff that you are doing anyway… you could make your note a lot shorter and just refer to that document [the automated Summary] … I like the idea that you can explore. Clinical genetics now is limited on time." [Tester 5]*

The DDSS also helped testers learn about diseases and associated findings with which they are less familiar.

> *"It's nice because it helps guide me… it's very easier for me to realize that Prader–Willi is associated with narcolepsy...." [Tester 3]*

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted August 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

The RoR report used by each tester in each scenario (n=9 RoR reports) includes a detailed prognosis table[22] that testers valued highly for being standardized and for its ability to communicate complex genetic information to patients and other clinicians.

> *"…The report is a great idea for highlighting why you think it's [the care instructions] important, [in] a standard format… The average primary care physician that gets the genetic testing reports, says I don't know what this means at all. I think this [the Prognosis Table] is a step towards making it more understandable." [Tester 2]*

Each of these three testers also exhibited learning and familiarity with GPACSS as they progressed through the three scenarios; appreciating the DDSS assistance as the cases increased in complexity.

> *"It takes it (clinical diagnosis and diagnostic thinking) to a higher level". [Tester 2]*

These testers also expressed readiness to adopt the tool in clinical practice.

> *"I would use it most of the time. To me, this is the frontier of genomic medicine and I look at my role as not only taking care of a patient, but figuring out how we make genomics part of everyday medical practice. The useful things in the chart, genetics people can now get to right away. [Tester 2]*

> *I think the interface is really good, in that you have that ability to explore those variants that may or may not make it on the reports that we get now, so you can drill deeper if you want. [Tester 5]*

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted August 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

One tester (pediatric geneticist) also suggested that GPACSS could help them as a differential diagnosis training tool for medical students and genetics residents.

Acceptability and fit beyond the differential diagnosis process

The additional testers selected for ability to comment on fit and acceptability in other areas of genetics (n=3) expressed GPACSS fit especially well with in-house sequencing laboratories. Because these testers interact with genomic data differently than the other testers, they raised the question of how GPACSS fits with their clinical genetic testing workflow in which labs provide a report with variants labeled as to pathogenicity and association with a condition, implying a clinical diagnosis. These testers, however, also identified value of GPACSS in genomic testing situations in which there is greater uncertainty as to the diagnosis or where flagged findings and the usefulness ranking would allow clinicians to draw on other information to make the diagnosis. Furthermore, these testers hypothesized that the ability to periodically re-analyze the VCF in minutes could cost-effectively improve the diagnosis rate over time.

The RoR report was also noted as an improvement over current laboratory reports by these testers. Some questioned if the prognosis table was “too detailed”, but after being apprised of the positive response to this table from prior testing,[19-22] they felt the report appropriately addressed an unmet clinical need.

> *“We've already done all that [annotated variants and associated it with a condition]… then just uploading it in [GPACSS]. To be able to have this report output, I think probably that's where the most utility would be”. [Tester 4]*

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted August 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

DISCUSSION

This study reports on the adaptation of a DDSS tool to a real-world clinical environment and testing within the clinical workflow for perceived need, acceptability, and workflow fit. Such assessment is critical if CDS are to fulfil the promise of standardizing and improving care.[1,4,5,8]

Technical issues expressed by GPACSS testers related to the interface and interaction of CDS design were minor and fixable; as were identified issues with design layout, and have already been instituted or planned. Despite initial concern for excess clicks and cognitive load, testers acknowledged the clicks as necessary to the genetic diagnosis process, and that the training video took a less efficient process for didactic purposes. Testers said that the cognitive load to determine age of onset was no different with or without the DDSS. Other issues related to terminology and usability may be solved through a combination of training, added info buttons, and experience with the tool.

GPACSS fits with the new CDS heuristic proposed by Miller et al[5] of “integration into real-time workflow,” as all testers identified ways GPACSS added such value and fit. Furthermore, according to Rogers’ Diffusion of Innovations model, organizations must evaluate innovations for fit to meet a perceived need and be an acceptable solution at the organization level.[28,29] All testers noted ways GPACSS filled multiple needs in the diagnostic process; therefore, GPACSS was found overall by testers to have good fit for organizational implementation regardless of individual tester issues and suggestions for technical improvements. Workflow fit was highest among core testers

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted August 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

involved in differential diagnosis of genetic conditions; yet other opportunities for workflow fit were described by the additional testers and included report generating capability to support communication where in-house sequencing is performed. Such workflow fit is essential for adoption and future implementation by individual clinicians and healthcare systems.

LIMITATIONS

To facilitate user testing of GPACSS in the context of clinical workflow prior to full integration and implementation, simulations of the real-world were required. First, this study used the Logica EHR simulation (Table 1); thus, the benefits or drawbacks GPACSS in a production EHR could not be directly observed. Similarly, the lack of full annotations for the causal variants in the variant table for the simulated patients did not permit full assessment of the value of the DDSS in variant interpretation by testers. Finally, the use of generic cTAKES NLP provided a minimum level of identification of findings to expect, but were not sufficient to guide how NLP improvements could improve recognition. However, even this basic NLP functionality was sufficient to return key terms to support differential diagnosis generation. These simulations were a necessary first step and provide data to further improve implementation potential of GPACSS. Beta testing within organizations actual EHRs, with real patient results, and NLP improvements will be necessary to fully assess individual user-level and organizational-level facilitators and barriers to use and implementation.

Additionally, the 4-minute training video was created to provide enough instruction to facilitate user testing only, and was not as detailed as a training tool necessary for

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted August 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

implementation. This resulted in testers noting gaps or requesting additions that were already in GPACSS and would be apparent to users receiving a detailed orientation to the tool. Results from this user testing provides necessary information to include in training and ongoing reference materials for future implementation in organizations.

CONCLUSIONS

Automated chart review combined with a DDSS can help improve the efficiency and consistency of genetic diagnosis. The current study surfaced some straightforward ways to improve the usability of such a system, many of which have already been implemented. Overall, tester responses suggest the GPACSS is usable based on technical CDS design criteria, addressed perceived clinical need, and most importantly, has good fit within the real-world clinical workflow of genetic testing and diagnosis process. Further development could improve implementation potential and fulfil the promise of care standardization and improved patient outcomes.

COMPETING INTERESTS

LKF and MMS are employed by SimulConsult and have an ownership stake.

All other authors declare no competing interests

FUNDING

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted August 24, 2020. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

This study was supported by the National Human Genome Research Institute of the National Institutes of Health under Award Number 1R43HG010322-01 (principal investigator: MMS). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted August 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Figure Legends:

Figure A: Architecture of the Genotype-Phenotype Archiving and Communication System with SimulConsult (GPACSS)

The key components are the coordination / archiving system (blue), the DDSS (green) and the NLP (yellow).

Figure B: Flagged findings with EHR text display for DDSS

A finding having a flag icon indicates that information was found in the EHR. Clicking the flag shows the various mentions of the flagged finding.

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted August 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

REFERENCES

1. Miller K, Mosby D, Capan M, et al. Interface, information, interaction: a narrative review of design and functional requirements for clinical decision support. *Journal of the American Medical Informatics Association* 2017;25(5):585-92.

2. Magrabi F, Ammenwerth E, McNair JB, et al. Artificial Intelligence in Clinical Decision Support: Challenges for Evaluating AI and Practical Implications. *Yearbook of medical informatics* 2019;28(1):128-34.

3. Bright TJ, Wong A, Dhurjati R, et al. Effect of clinical decision-support systems: a systematic review. *Ann Intern Med* 2012;157(1):29-43.

4. Moxey A, Robertson J, Newby D, et al. Computerized clinical decision support for prescribing: provision does not guarantee uptake. *J Am Med Inform Assoc* 2010;17(1):25-33.

5. Miller K, Capan M, Weldon D, et al. The design of decisions: Matching clinical decision support recommendations to Nielsen's design heuristics. *International journal of medical informatics* 2018;117:19-25.

6. Moja L, Liberati EG, Galuppo L, et al. Barriers and facilitators to the uptake of computerized clinical decision support systems in specialty hospitals: protocol for a qualitative cross-sectional study. *Implementation science : IS* 2014;9:105.

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted August 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

7. Liberati EG, Ruggiero F, Galuppo L, et al. What hinders the uptake of computerized decision support systems in hospitals? A qualitative study and framework for implementation. *Implementation science : IS* 2017;12(1):113.

8. Sutton RT, Pincock D, Baumgart DC, et al. An overview of clinical decision support systems: benefits, risks, and strategies for success. *NPJ digital medicine* 2020;3:17.

9. Williams MS. Perspectives on what is needed to implement genomic medicine. *Molecular genetics & genomic medicine* 2015;3(3):155-9.

10. Shyr C, Kushniruk A, van Karnebeek CD, et al. Dynamic software design for clinical exome and genome analyses: insights from bioinformaticians, clinical geneticists, and genetic counselors. *J Am Med Inform Assoc* 2016;23(2):257-68.

11. Segal MM, Williams MS, Gropman AL, et al. Evidence-based decision support for neurological diagnosis reduces errors and unnecessary workup. *Journal of child neurology* 2014;29(4):487-92.

12. Welch BM, Kawamoto K. Clinical decision support for genetically guided personalized medicine: a systematic review. *J Am Med Inform Assoc* 2013;20(2):388-400.

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted August 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

13. Grebe TA, Khushf G, Chen M, et al. The interface of genomic information with the electronic health record: a points to consider statement of the American College of Medical Genetics and Genomics (ACMG). *Genetics in Medicine* 2020.

14. Segal MM, George R, Waltman P, et al. Clinician-centric diagnosis of rare genetic diseases: performance of a gene pertinence metric in decision support for clinicians. *Orphanet J. Rare Diseases* 2020 (in press).

15. Segal MM, Abdellateef M, El-Hattab AW, et al. Clinical pertinence metric enables hypothesis-independent genome-phenome analysis for neurologic diagnosis. *Journal of child neurology* 2015;30(7):881-8.

16. Issa MY, Chechlacz Z, Stanley V, et al. Molecular diagnosis in recessive pediatric neurogenetic disease can help reduce disease recurrence in families. *BMC medical genomics* 2020;13(1):68.

17. Kassirer JP. Teaching clinical medicine by iterative hypothesis testing. Let's preach what we practice. *The New England journal of medicine* 1983;309(15):921-3.

18. Eddy DM, Clanton CH. The art of diagnosis: solving the clinicopathological exercise. *The New England journal of medicine* 1982;306(21):1263-8.

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted August 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

19. Williams MS, Williams JL, Rahm A, et al. Evaluating the Impact of a User-Friendly Report of Genetic Testing Results for Parents of Children with Genetic Disorders. Washington, DC: Patient-Centered Outcomes Research Institute (PCORI), 2018.

20. Williams JL, Rahm AK, Zallen DT, et al. Impact of a Patient-Facing Enhanced Genomic Results Report to Improve Understanding, Engagement, and Communication. *Journal of genetic counseling* 2017.

21. Williams JL, Rahm AK, Stuckey H, et al. Enhancing genomic laboratory reports: A qualitative analysis of provider review. *American journal of medical genetics. Part A* 2016;170(5):1134-41.

22. Segal MM, Rahm AK, Hulse NC, et al. Experience with Integrating Diagnostic Decision Support Software with Electronic Health Records: Benefits versus Risks of Information Sharing. *EGEMS (Washington, DC)* 2017;5(1):23.

23. Savova GK, Masanz JJ, Ogren PV, et al. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. *J Am Med Inform Assoc* 2010;17(5):507-13.

24. Segal MM. 2004. Systems and methods for diagnosing medical conditions US patent 6,754,655.

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted August 24, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

25. Khajouei R, Zahiri Esfahani M, Jahani Y. Comparison of heuristic and cognitive walkthrough usability evaluation methods for evaluating health information systems. *J Am Med Inform Assoc* 2017;24(e1):e55-e60.

26. Khajouei R, Farahani F. A combination of two methods for evaluating the usability of a hospital information system. *BMC Med Inform Decis Mak* 2020;20(1):84.

27. Nielsen J. Enhancing the explanatory power of usability heuristics. . *Proc. ACM CHI'94 Conf. (Boston, MA, April 24-28)* *https://www.nngroup.com/articles/ten-usability-heuristics/* 1994:152-58.

28. Rogers EM. Innovation in Organizations. Diffusion of Innovations. 5th ed. New York, NY: Simon & Schuster, Inc, 2003:402 - 35.

29. Mittman B. Implementation Science in Health Care. In: Brownson RC CG, Proctor EK, ed. Dissemination and Implementation Research in Health. New York: Oxford University Press, 2012:400-18.

(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license.

SimulConsult EHR 1

Initial | Diagnose | Profile | Assess | Tips

Differential | **Add findings** | Add tests | Phenotype 2 1 | Genotype

- Marfan syndrome, classical
- XYY syndrome (double Y...
- Marfan syndrome, early ...
- Fragile X syndrome, males
- Simpson-Golabi-Behmel ...
- Sotos-Dodge syndrome
- Geleophysic dysplasia 2, ...
- NF1: Neurofibromatosis 1
- NS1: Noonan syndrome, ...
- Centronuclear myopathy...
- Trisomy 21 (Down syndr...
- Lujan-Fryns syndrome
- Intellectual development...
- Homocystinuria due to c...
- NS5: Noonan syndrome, ...
- NS4: Noonan syndrome, ...
- Loeys-Dietz syndrome 5,...
- FG syndrome
- Centronuclear myopathy...
- Neurofibromatosis-Noo...
- Hypomelanosis of Ito

Top findings ranked by usefulness in narrowing the differential diagnosis

All findings shown | Only flagged input findings shown

Presence	Findings ranked by usefulness
?	Myopia, severe
?	Digits: slender fingers and toes
?	Palpebral fissure downslanted
?	
?	
?	
?	
?	
?	
?	
?	
?	
?	
?	
?	
?	Dolichocephaly
?	Arm span exceeds height
?	Reduced upper/lower segment ratio
?	Digits: fingers long

The finding:

Myopia, severe

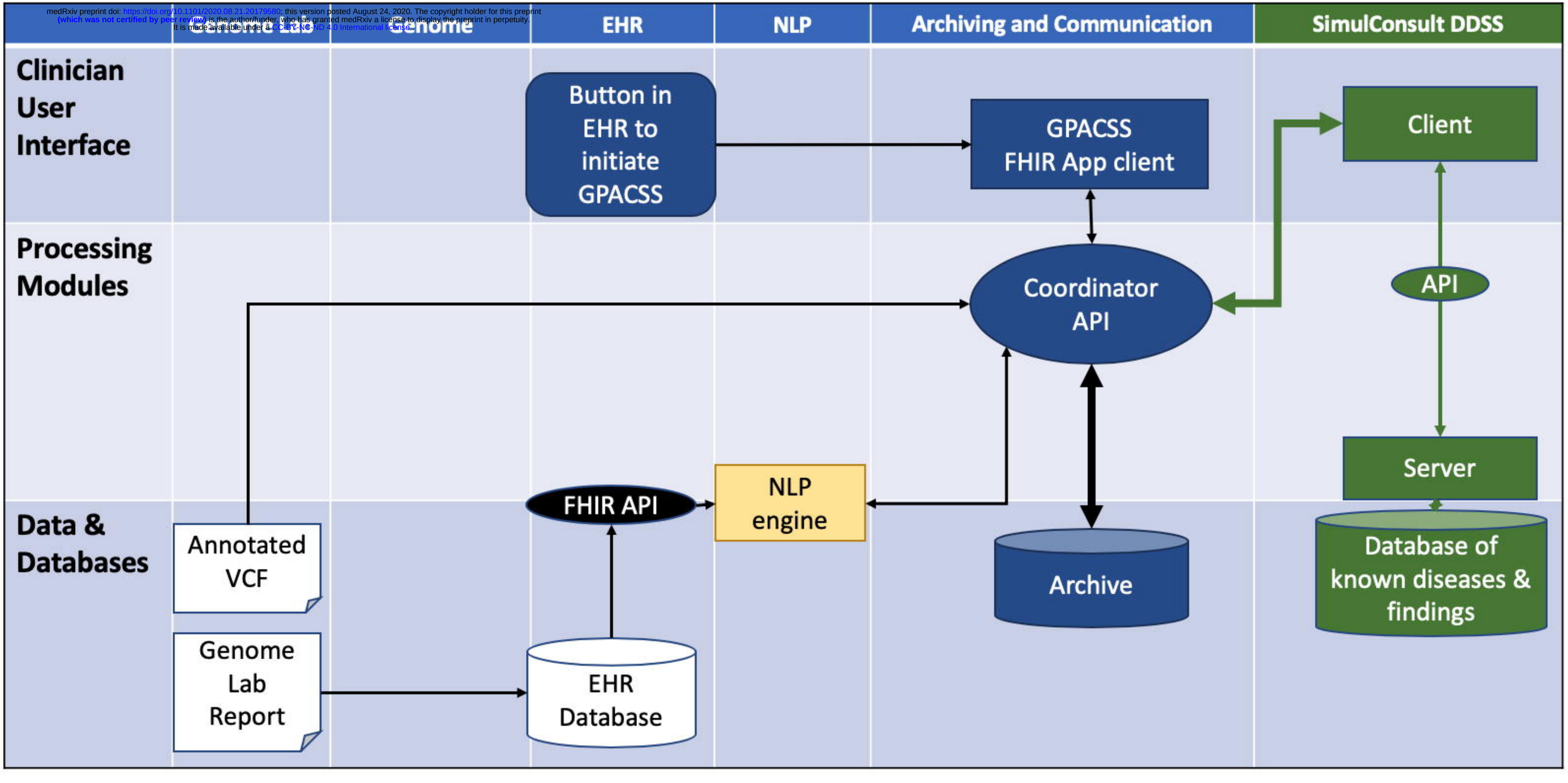
was flagged in the following contexts:

On 2000-05-31 (age 8 years) *Dr. Rodrigo Mills* documented in the EHR

"Eyes : **Wears glasses , Hx of Myopia.** Chest :"

On 2002-10-24 (age 10 years) *Dr. Rodrigo Mills* documented in the EHR

"Wears glasses. **Hx of myopia.** HEENT :"

NOTE: Information used to flag a finding is processed on the local computer and is not sent to the SimulConsult server.

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted August 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC BY-NC-ND 4.0 International license.